4SL4 Assignment 2

Aaron Pinto
400190637
pintoa9

I used 5-fold cross-validation, because that's what suggested here https://scikit-learn.org/stable/modules/cross_validation.html.

Here are the K-fold cross-validation errors for the base model, for each feature, and the subset of selected features.

		Cross-validation Errors												
K	Subset of selected features	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
1	LSTAT	70.37414	74.3865	63.7868	80.4375	68.60608	44.25055	73.14208	78.05871	70.21044	64.27843	60.42195	73.31079	38.6745
2	LSTAT, RM	38.40016	39.12265	38.639	37.08678	38.684	31.47249	38.17896	37.13206	38.79695	38.12066	33.67586	38.45224	
3	LSTAT, RM, PTRATIO	30.92364	31.65331	31.51299	30.72503	31.54114		31.55344	30.84132	31.28542	30.65951	28.56417	30.98923	
4	LSTAT, RM, PTRATIO, DIS	28.28051	28.69884	28.71289	28.17901	28.55511		28.41385	27.61117	28.59951	28.45782		28.13951	
5	LSTAT, RM, PTRATIO, DIS, NOX	27.108	27.68883	27.32087	27.42743	26.1118		27.58158		27.61719	27.14709		27.03184	
6	LSTAT, RM, PTRATIO, DIS, NOX, B	25.83525	26.13661	26.22155	25.87564			26.10659		26.07171	26.11502		25.78344	
7	LSTAT, RM, PTRATIO, DIS, NOX, B, RAD	25.67763	25.76747	25.88962	25.64228			25.79509		25.60791	25.81449			
8	LSTAT, RM, PTRATIO, DIS, NOX, B, RAD, TAX	25.06611	25.67212	25.71526	25.42869			25.61998			25.04905			
9	LSTAT, RM, PTRATIO, DIS, NOX, B, RAD, TAX, CRIM	24.50314	24.92451	25.16097	24.959			25.06346						
10	LSTAT, RM, PTRATIO, DIS, NOX, B, RAD, TAX, CRIM, ZN		24.28446	24.60889	24.44723			24.53131						
11	LSTAT, RM, PTRATIO, DIS, NOX, B, RAD, TAX, CRIM, ZN, CHAS			24.34195	24.18913			24.28958						
12	LSTAT, RM, PTRATIO, DIS, NOX, B, RAD, TAX, CRIM, ZN, CHAS, AGE			24.26006				24.21342						
13	LSTAT, RM, PTRATIO, DIS, NOX, B, RAD, TAX, CRIM, ZN, CHAS, AGE, INDUS			24.2872										

Here are the cross-validation errors for all the models I considered.

	Cross-validation errors		
K	Base model	Polynomial degree 2	Logarithmic
1	38.6745025	30.49227149	28.31625255
2	31.47249464	22.42679689	26.41025999
3	28.56417233	19.44032441	24.09370554
4	27.6111696	17.03528062	22.25118965
5	26.11180175	16.25876333	20.90459995
6	25.78344326	16.72007481	20.5260898
7	25.60791017	14.5497296	20.26092749
8	25.04904839	14.53722404	19.13069239
9	24.50313652	15.49021098	19.22547769
10	24.28445873	15.76654516	19.33998799
11	24.1891286	19.89075602	19.23379488

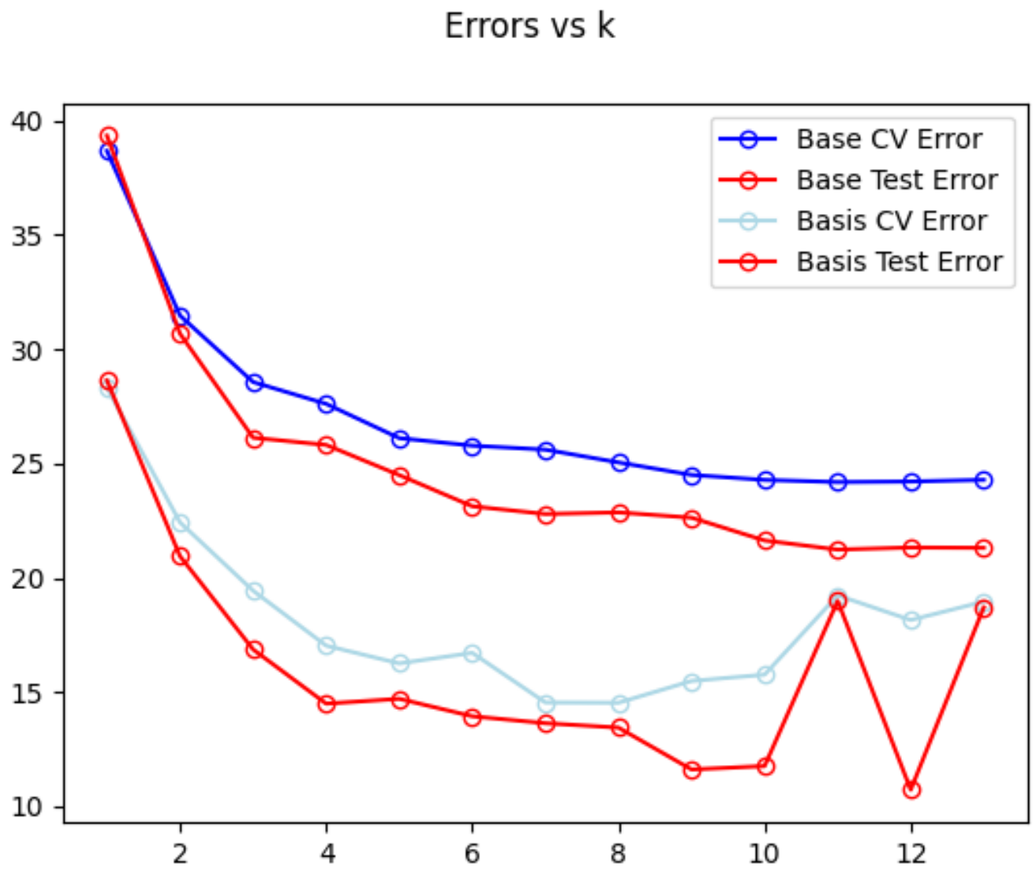
12	24.2134237	18.15598339	19.02009917
13	24.28719512	19.41591795	18.96132696

I compared the cross-validation errors between the polynomial and logarithmic models for each k, and then I used the cross-validation and test error from the model with the lower cross-validation score at that k. At k=1, 11, and 13, the logarithmic model had the lower cross-validation error. Otherwise, the polynomial model was lower.

K	Base model parameter vector	Best basis model parameter vector	Basis expansion
1	[34.12381541 -0.9254271]	[5.17963389e+09 -1.23761085e+01]	Logarithmic
2	[-0.74168906 -0.62849675 4.9577728]	[2.13927766e+01 2.13927766e+01 9.48875453e-02 -8.00222885e+00 2.13927766e+01 9.48875453e-02 -8.00222885e+00 1.06703108e-02 -2.08075573e-01 1.75290834e+00]	Polynomial of degree 2
3	[19.20907845 -0.55956188 4.28461418 -0.89685083]	[-4.99922407e+01 -4.99922407e+01 6.94436323e-01 8.05714295e+00 6.08212563e+00 -4.99922407e+01 6.94436323e-01 8.05714295e+00 6.08212563e+00 1.39277855e-02 -1.37900312e-01 -8.60770151e-02 9.48522033e-01 -1.24281879e+00 -1.07686931e-01]	Polynomial of degree 2
4	[25.71270154 -0.66156767 3.95202512 -0.94429952 -0.59508421]	[-5.24938413e+01 -5.24938413e+01 -2.31558525e-01 1.20491876e+01 6.98241773e+00 -6.22766491e+00 -5.24938413e+01 -2.31558525e-01 1.20491876e+01 6.98241773e+00 -6.22766491e+00 3.10954651e-02 -1.06012535e-01 -6.44960757e-02 2.15484651e-01 7.18132801e-02 -1.28039540e+00 1.09007359e+00 -1.63485697e-01 5.31076942e-02 1.87254881e-01]	Polynomial of degree 2
5	[38.27736569 -0.5774377 3.98635901 -0.98351286 -1.24929695 -19.26063043]	[-5.58222658e+01 -5.58222658e+01 6.64188104e-01 1.67520198e+01 4.61739932e+00 -4.93917976e+00 2.27415734e+01 -5.58222658e+01 6.64188104e-01 1.67520198e+01 4.61739932e+00 -4.93917976e+00 2.27415734e+01 3.27309688e-02 -5.07525159e-02 -9.56472587e-02 1.14423633e-01 -2.00938902e+00 8.38413380e-02 -1.16535845e+00 4.41587214e-01 -1.70289976e+01 -7.84455985e-02 1.35526181e-01 1.75727635e+00 2.26736765e-01 -2.08663530e-02 2.97766567e+01]	Polynomial of degree 2
6	[3.30487222e+01 -5.49036061e-01 4.08086588e+00 -9.72671230e-01 -1.21606658e+00 -1.72191795e+01 7.89036003e-03]	[-1.76892257e+08 8.84460087e+07 1.07727958e+00 1.85933419e+01 5.42656548e+00 -4.34417263e+00 4.39121760e+01 7.65565881e-02 8.84460087e+07 1.07745233e+00 1.85937816e+01 5.42661503e+00 -4.34392883e+00 4.39120799e+01 7.65549860e-02 2.87794604e-02 -9.22618099e-02 -9.98835708e-02 1.08097228e-01 -2.18991746e+00 -7.46796391e-04 3.00145508e-02 -1.18265289e+00 4.61275136e-01 -1.79390750e+01 -5.23627715e-03 -8.29852847e-02 1.08145101e-01 1.17204166e+00 -2.28714546e-03 2.31409288e-01 -8.44829323e-01 -1.19190602e-03 2.41265950e+01 -3.82022828e-02 -6.69658184e-05]	Polynomial of degree 2
7	[3.67505934e+01 -5.62048721e-01 3.95015064e+00 -1.08679444e+00 -1.23983743e+00 -2.06577682e+01 9.43509229e-03 8.69408163e-02]	[1.95828164e+08 -9.79141651e+07 5.76658960e-01 1.76396442e+01 2.54117467e+00 -4.48999787e+00 2.32275123e+01 6.43852495e-02 5.46257585e-01 -9.79141651e+07 5.76885009e-01 1.76395129e+01 2.54077033e+00 -4.49003219e+00 2.32277188e+01 6.44791431e-02 5.46266436e-01 2.27708280e-02 -1.69060384e-01 -1.06601181e-02 8.23592092e-02 -7.73347751e-01 -1.64076710e-03 -3.97309101e-02 -2.09120761e-01 -8.62748236e-01 3.21420397e-01 -1.38186459e+01 -7.10604077e-03 -1.25176276e-01 -6.71745409e-02 1.75391288e-01 -4.39191641e-01 3.63561418e-03 9.18163983e-02 2.47709018e-01 5.27894065e+00 -6.91093970e-03 -1.13196635e-01 3.47353092e+01 -5.39600039e-02 -2.96878127e-01 -8.00192356e-05 -1.60553213e-03 -1.40695888e-02]	Polynomial of degree 2
8	[3.81477632e+01 -5.58416224e-01 3.80850810e+00 -1.03477933e+00 -1.21148179e+00 -1.76051812e+01 9.10131536e-03 2.67441240e-01 -1.20483834e-02]	[3.78108850e+07 -1.89054858e+07 4.31971968e-01 1.57274731e+01 3.18039180e-01 -3.62492651e+00 1.51540660e+01 5.25976780e-02 9.01305419e-01 -7.55858669e-02 -1.89054858e+07 4.31562404e-01 1.57278887e+01 3.17944680e-01 -3.62510661e+00 1.51541033e+01 5.26967343e-02 9.01226481e-01 -7.55602310e-02 2.08983271e-02 -1.85367416e-01 1.58737925e-02 8.71852694e-02 -1.16291723e-01 -1.80829526e-03 -1.81319850e-02 -1.50452857e-03 -1.44885428e-01 -6.81677389e-01 2.92135628e-01 -1.19078011e+01 -8.20848159e-03 -1.22624889e-01 -1.91396568e-03 -2.37839412e-02 1.33824236e-01 -9.61564586e-01 6.02766406e-03 8.86529949e-02 3.48146819e-03 2.21088528e-01 4.82491399e+00 -8.40495899e-03 -1.95017829e-01 2.95998715e-03 7.49815736e+00 -2.67506388e-02 -2.82482434e+00 1.38139616e-01 -8.86656344e-05 -1.18539878e-03 -5.41154295e-05 -1.65730414e-01 7.23945850e-03 1.66706741e-05]	Polynomial of degree 2
9	[3.90303347e+01 -5.31400175e-01 3.80662732e+00 -1.05462489e+00 -1.24977673e+00 -1.85454409e+01 8.22796712e-03 3.23043850e-01 -1.20967438e-02 -1.01845026e-01]	[-1.20223018e+07 6.01110503e+06 4.16606389e-01 1.51492130e+01 8.12448575e-01 -4.15922539e+00 1.72775087e+01 5.49587687e-02 1.18346825e+00 -8.34975325e-02 -3.55067899e-01 6.01110503e+06 4.16996578e-01 1.51489412e+01 8.12498085e-01 -4.15922144e+00 1.72775279e+01 5.49433464e-02 1.18352693e+00 -8.35065864e-02 -3.55055120e-01 1.52636494e-02 -2.53689594e-01 2.33928675e-02 1.06071907e-01 4.41087815e-01 -1.40654495e-03 -1.09265526e-02 -1.83854940e-03 1.09747270e-02 -2.51900308e-01 -6.54460737e-01 3.98594816e-01 -8.64704611e+00 -4.90859372e-03 -1.40274638e-01 -4.17107274e-03 1.34406175e-01 -2.12062593e-02 1.02962326e-01 -7.37501107e-01 2.49534572e-03 6.99172646e-02 4.44344227e-03 -2.29775305e-01 2.11283796e-01 4.51608516e+00 -4.95819922e-03 -2.28769298e-01 2.56304361e-03 -4.75696360e-02 2.15847383e+00 -5.66751473e-02 -1.47660966e+00 6.87939702e-02 -1.15993958e+00 -7.54408538e-05 -1.94072993e-03 4.58015129e-05 -2.83878297e-04 -1.58511938e-01 6.24258514e-03 -2.88499600e-01 3.66703607e-05 1.79663437e-02 8.80147098e-04]	Polynomial of degree 2
10	[3.83568262e+01 -5.33923027e-01 3.65680335e+00 -9.23656925e-01 -1.52643436e+00 -1.74970452e+01 8.12770928e-	[2.82095751e+07 -1.41048410e+07 4.50847332e-01 1.59100591e+01 1.73877592e+00 -3.62288957e+00 -3.62291540e+00 5.54504664e-02 1.38340875e+00 -6.63278341e-02 -2.10419219e+00 -1.84732697e-01 -1.41048410e+07 4.50747846e-01 1.59090517e+01 1.73871110e+00 -3.62299056e+00 -3.62297369e+00 5.53173695e-02 1.38339941e+00 -6.63099086e-02 -2.10423278e+00 -1.84759248e-01 1.35077044e-02 -2.49264405e-01 2.99686570e-02 9.75177880e-02	Polynomial of degree 2

	03 3.39937084e-01 -1.41984216e-02 -1.09537534e-01 4.23061195e-02]	7.70023595e-01 -1.67658411e-03 3.73697579e-03 -2.76639165e-03 1.02648103e-02 -4.04591236e-04 -2.17149190e-01 -7.17874569e-01 3.29469611e-01 -7.63163291e+00 -5.68141532e-03 -4.57281815e-02 -9.29592736e-03 1.37669301e-01 -1.88536360e-03 -4.86198170e-02 2.34420886e-02 -2.20993669e+00 4.27377771e-03 1.70746117e-02 4.48260678e-03 -1.10847679e-01 9.44661908e-04 3.57517578e-01 9.55082674e+00 -1.04013660e-02 -1.60146972e-01 -7.30339903e-04 -8.55431828e-02 -1.32383481e-02 8.63703126e+00 -3.89141757e-02 -2.75278710e+00 1.61991595e-01 -1.00297595e+00 -2.12924512e-01 -8.41245055e-05 -1.38471834e-03 -6.95325434e-06 -3.81343067e-04 9.18619800e-04 -1.39098243e-01 5.87948528e-03 -3.72103983e-01 -1.17468562e-02 1.95149332e-05 2.26332708e-02 5.09489328e-04 5.49484510e-04 2.90294235e-01 8.75153208e-04]	
11	[3.81511459e+01 -5.35284690e-01 3.56375105e+00 -8.86823804e-01 -1.48937016e+00 -1.79448580e+01 7.65944186e-03 3.22989562e-01 -1.33477769e-02 -1.03827907e-01 4.29268833e-02 2.79831373e+00]	[8.42950738e+09 -9.77299406e+00 1.32090147e+01 -1.34411280e+01 -6.58771224e+00 -1.18956320e+01 9.58594270e-01 2.10760623e+00 -5.13578782e+00 5.49324275e-02 6.21298628e-03 1.20856634e-01]	Logarithmic
12	[3.83495297e+01 -5.42548594e-01 3.52621916e+00 -8.91800638e-01 -1.46003026e+00 -1.85107809e+01 7.56260772e-03 3.25538996e-01 -1.34035466e-02 -1.04034796e-01 4.36239637e-02 2.77082958e+00 6.56651467e-03]	[2.01925686e+07 -1.00963349e+07 2.83140952e-01 1.16394763e+01 5.74655286e-01 -2.09967241e+00 2.47185204e+01 5.71270764e-02 2.71745212e+00 -1.47973875e-01 -4.69325667e+00 2.09344981e-01 1.22711412e+01 5.02986559e-01 -1.00963349e+07 2.85323882e-01 1.16372082e+01 5.74353128e-01 -2.09986026e+00 2.47189311e+01 5.70377077e-02 2.71752957e+00 -1.48111337e-01 -4.69333947e+00 2.09331373e-01 1.22711017e+01 5.03011203e-01 9.89804968e-03 -2.30293723e-01 3.68840016e-02 1.13726670e-01 1.25556616e+00 -6.78422541e-04 -1.29484959e-02 -1.83381676e-03 2.24107835e-02 -5.28849883e-03 -9.57105176e-02 -7.94890127e-03 1.42292162e-01 -4.85132488e-01 4.54963246e-02 -1.24732467e+00 -6.36766874e-03 -3.37911897e-01 1.99053017e-03 1.40208091e-01 -2.99443031e-02 -3.32872981e+00 -4.51320186e-02 5.86550622e-02 -1.36812263e-01 -6.69989356e+00 3.21662147e-03 -7.43831573e-02 6.99060468e-03 1.40524044e-01 1.23708742e-04 -8.98496321e-01 1.22007914e-05 4.34679545e-01 1.14324292e+01 -6.75053592e-03 -5.57726627e-02 -3.21198616e-03 -9.68552540e-02 -2.76594698e-02 1.73084749e+00 -2.02294576e-02 -2.94204518e+01 -8.24078452e-05 -4.28673699e+00 2.83162226e-01 -1.03683216e+00 -7.46390182e-01 -3.69630689e+01 -2.97972640e-01 -5.28283417e-05 -1.32782396e-03 1.95186585e-05 -3.15676443e-04 3.72739509e-04 2.13714374e-02 -7.32631655e-04 -1.47149656e-01 6.08014959e-03 -6.02439814e-01 -1.04284083e-02 -2.38754743e-01 2.16171343e-02 5.29810786e-05 3.10172276e-02 4.80536371e-04 1.28823668e-02 -8.03611983e-04 1.80145283e-03 1.25035830e-01 2.29853637e+00 -3.31305785e-03 5.75524755e-04 -7.37106028e-02 2.10108934e-04 1.22710877e+01 5.74665537e-02 -4.89033991e-04]	Polynomial of degree 2
13	[3.84595290e+01 -5.44007615e-01 3.53748901e+00 -8.98333476e-01 -1.44710354e+00 -1.88260874e+01 7.59360665e-03 3.30633998e-01 -1.38600746e-02 -1.03581852e-01 4.40837095e-02 2.74529904e+00 6.50038618e-03 1.88719049e-02]	[8.14010323e+09 -9.99946777e+00 1.20792195e+01 -1.31748436e+01 -6.35030807e+00 -1.23594781e+01 9.45519599e-01 2.08896940e+00 -4.85709792e+00 1.07690695e-01 -9.29128685e-04 1.19983963e-01 9.46044444e-01 -6.12463677e-01]	Logarithmic

Here is a plot of the cross-validation errors and test errors for the base model and for the best (based on K-fold cross-validation) basis expansion model.



The cross-validation error is almost always larger than the test error for both models. Except for at k=1, the test error on the entire subset is always less than the cross-validation error for the best basis model and for the base model. This is because k-fold cross validation tends to overestimate generalization error. But as the error of the model decreases with increasing training sample size as we add more features, the substitute model on average has (slightly) higher true generalization error than the model trained on the whole data set - which is the model whose error is approximated by the cross validation. This relation is consistent for both the base and the basis expansion models.

For the base model, the model at k=11 had the smallest cross-validation error. That model also has the smallest test error. For the model with basis expansion, the model at k=8 had the smallest cross-validation error. The model at k=11 had the smallest test error.